

Speech intelligibility in rooms: Effect of prior listening exposure interacts with room acoustics

Pavel Zahorik^{a)}

Department of Otolaryngology and Communicative Disorders, University of Louisville, Louisville, Kentucky 40292, USA

Eugene J. Brandewie

Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455, USA

(Received 22 September 2015; revised 29 April 2016; accepted 9 June 2016; published online 6 July 2016)

There is now converging evidence that a brief period of prior listening exposure to a reverberant room can influence speech understanding in that environment. Although the effect appears to depend critically on the amplitude modulation characteristic of the speech signal reaching the ear, the extent to which the effect may be influenced by room acoustics has not been thoroughly evaluated. This study seeks to fill this gap in knowledge by testing the effect of prior listening exposure or listening context on speech understanding in five different simulated sound fields, ranging from anechoic space to a room with broadband reverberation time (T60) of approximately 3 s. Although substantial individual variability in the effect was observed and quantified, the context effect was, on average, strongly room dependent. At threshold, the effect was minimal in anechoic space, increased to a maximum of 3 dB on average in moderate reverberation (T60 = 1 s), and returned to minimal levels again in high reverberation. This interaction suggests that the functional effects of prior listening exposure may be limited to sound fields with moderate reverberation ($0.4 \leq T60 \leq 1$ s). © 2016 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4954723>]

[EAS]

Pages: 74–86

I. INTRODUCTION

It has long been known that speech understanding is degraded by reverberation (Knudsen, 1929). The degradation stems primarily from temporal distortion of the speech signal caused by the reverberation (Bolt and MacDonald, 1949; Houtgast and Steeneken, 1985), and is known to scale with the amount of reverberation (Knudsen, 1929; Lochner and Burger, 1964). Because most everyday communication situations involve sound transmission within a reverberant sound field, it is critically important to understand how and by what mechanisms speech understanding is impacted by reverberation. This is especially true, given that the negative effects of reverberation on speech understanding are exacerbated both by background noise (Knudsen, 1929; Nabelek and Mason, 1981)—another ubiquitous property of everyday listening environments—and by hearing loss (Gelfand and Hochberg, 1976), where poor performance in reverberation is the most frequent complaint given by hearing aid users (Johnson *et al.*, 2010). Given these challenges, it is perhaps remarkable that for individuals with normal hearing, few communication problems are encountered in everyday reverberant environments. This suggests that processing in the normally functioning auditory system must effectively counteract the deleterious effects of reflected sound and reverberation even though, acoustically, these effects are clearly measurable and specific to a given listening

environment and the spatial configuration of components in the communication chain.

The auditory system has a number of mechanisms that can immediately assist with speech understanding in reverberation, including the binaural system (Moncur and Dirks, 1967) and mechanisms related to the precedence effect (see Litovsky *et al.*, 1999 and Brown *et al.*, 2015 for reviews). Beyond these immediate effects, there is now emerging evidence that prior listening exposure to reverberation can provide an environmental listening context that renders speech as perceptually less reverberant (Watkins, 2005a,b) and can result in objective improvements in speech intelligibility (Brandewie and Zahorik, 2010). This suggests that the processing of sound in reverberant space within the normally functioning auditory system may involve processes more complicated than previously thought. The goal of the present study is to determine the extent to which this environmental context effect depends on the acoustics of the listening space.

Watkins (2005a,b) was the first to demonstrate an effect of listening context on speech perception in reverberation. He used target speech signals on an 11-point continuum from “sir” to “stir” embedded in a carrier phrase, and noted the point at which the speech percept changed from “sir” to “stir”—a categorical perception task. When both target and carrier phrase were presented in minimal reverberation, the change point was near the center of the continuum. When the target was presented in moderate reverberation, but the carrier phrase remained in minimal reverberation, the change point shifted toward “sir.” This can be explained by reverberant energy filling in the temporal gap following the stop

^{a)}Also at: Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY and Heuser Hearing Institute, Louisville, KY. Electronic mail: pavel.zahorik@louisville.edu

consonant in “stir,” causing it to be perceived as more like “sir.” When both the target and the carrier phrase were presented in reverberation, the change point shifted back to where it was observed when both target and carrier were presented in minimal reverberation. This suggests that the reverberant carrier phrase provides contextual information that allows the auditory system to compensate for the effects of the reverberation on the target word. Watkins and his colleagues have interpreted this result as being consistent with a type of high-level perceptual constancy, similar to other well-known perceptual constancies in vision, such as brightness constancy or color constancy. They have also demonstrated the effect with additional speech continua (Beeston *et al.*, 2014) and non-speech contexts (Watkins and Makin, 2007a,b), and have shown that the effect is driven primarily by the amplitude envelope of the speech signal reaching the listener (Watkins *et al.*, 2011). This latter result is appealing because of its potential links to the modulation transfer function concept, which forms the basis for standard methods of predicting speech intelligibility in rooms, such as the speech transmission index (STI; IEC-60268-16, 2003) and speech intelligibility index (SII; ANSI-S3.5-1997, 1997).

Brandewie and Zahorik (2010) reported context effects similar to those identified by Watkins and his colleagues, but using different methods. In their study, Brandewie and Zahorik (2010) compared speech reception thresholds (SRTs) using the coordinate response measure (CRM; Bolia *et al.*, 2000) in a background of spatially separate noise within a simulated reverberant room. Two different listening conditions were tested. In one condition, listeners were provided with consistent listening exposure to the same reverberant room, both within and across trials. In a second condition, consistent exposure to the room was disrupted by removing the CRM carrier phrase and changing the room from trial to trial. SRTs were found to be 2–3 dB lower, on average, in the consistent exposure condition. This suggests that consistent environmental listening context in a reverberant room can facilitate speech understanding. Similar context effects were not observed when the test room was anechoic, suggesting that the context effects are specific to reverberant sound fields. Additional work has demonstrated that the effect generalizes to highly heterogeneous sentence materials (Srinivasan and Zahorik, 2013), and is fully activated by ~1 s of listening exposure (Brandewie and Zahorik, 2013). The importance of the amplitude envelope in the room context effect has also been demonstrated using similar methods (Srinivasan and Zahorik, 2014).

Because Brandewie and Zahorik only evaluated the room context effect for a single, moderately reverberant room, and other related work has evaluated at most two sound fields (Watkins, 2005b; Srinivasan and Zahorik, 2014), it is important to determine the extent to which the effect may be room dependent. The current study was therefore designed to extend the work of Brandewie and Zahorik (2010) by testing context effects in five types of sound fields, ranging from anechoic to a highly reverberant room, using methods identical to those in Brandewie and Zahorik (2010). The differing reverberant sound fields will also help to better interpret the context effects by allowing intelligibility results

to be compared to those predicted by the STI. Because substantial individual variability was observed in Brandewie and Zahorik’s (2010) data ($n = 14$), the current study was also designed to better quantify this variability by testing many more listeners. The issue of individual variability is important to address because it has not been a focus of previous work on the effects of listening context in reverberation.

II. METHODS

A. Subjects

A total of 50 listeners (22 male, 28 female; age range 18.3–31.1 yr; median age 21.0 yr) participated in the experiment. All had normal hearing (pure tone thresholds ≤ 20 dB hearing level (HL) in both ears from 250 to 8000 Hz) and were fluent in English. Data from one listener were excluded because this listener only completed a portion of the experiment, and that portion resulted in poor psychometric function fits (see Sec. II E for details). All procedures involving human subjects were approved by the University of Louisville Institutional Review Board.

B. Stimuli

1. Sound field simulation

A virtual auditory space (VAS) technique identical to that used in Brandewie and Zahorik (2010) and described in detail in Zahorik (2009) was used to simulate sound field listening in five different environments. Briefly, this room simulation technique uses an image-model (Allen and Berkley, 1979) to precisely simulate early reflections and a statistical model to simulate late reverberant energy. The direct-path and early reflections are spatially rendered using non-individualized head-related transfer function (HRTF) measurements. The result of the simulation is an estimated binaural room impulse response (BRIR) that describes the transformation of sound between the source and the listeners’ ears in the simulated room. Zahorik (2010) found that this simulation technique produced BRIRs that were reasonable physical and perceptual approximations to those measured in a real office-sized rectangular room.

Five rooms were simulated in this experiment (R0–R4). All had identical dimensions (length, 5.7 m; width, 4.3 m; height, 2.6 m), but varied in the absorptive properties of the reflecting surfaces. The absorption coefficients for R2 were identical to those from Zahorik (2009), and were designed to approximate a moderately reverberant large office. R1 was less reverberant than R2, and R3 and R4 were more reverberant than R2. R0 was a simulated anechoic space, with all absorption coefficients set to one. Octave-band reverberation times (T60), clarity indices (C50), and interaural cross correlation (IACC) values for each simulated room are displayed in Fig. 1. T60 is a measure of the time it takes for reverberant sound to decay by 60 dB; C50 is a measure of the balance between early (0–50 ms) and late-arriving (50– ∞ ms) energy; and IACC is a measure of the relationship between the signals at the two ears (ISO-3382, 1997). No attempt was made to equalize sound levels across the three rooms. As a result, the

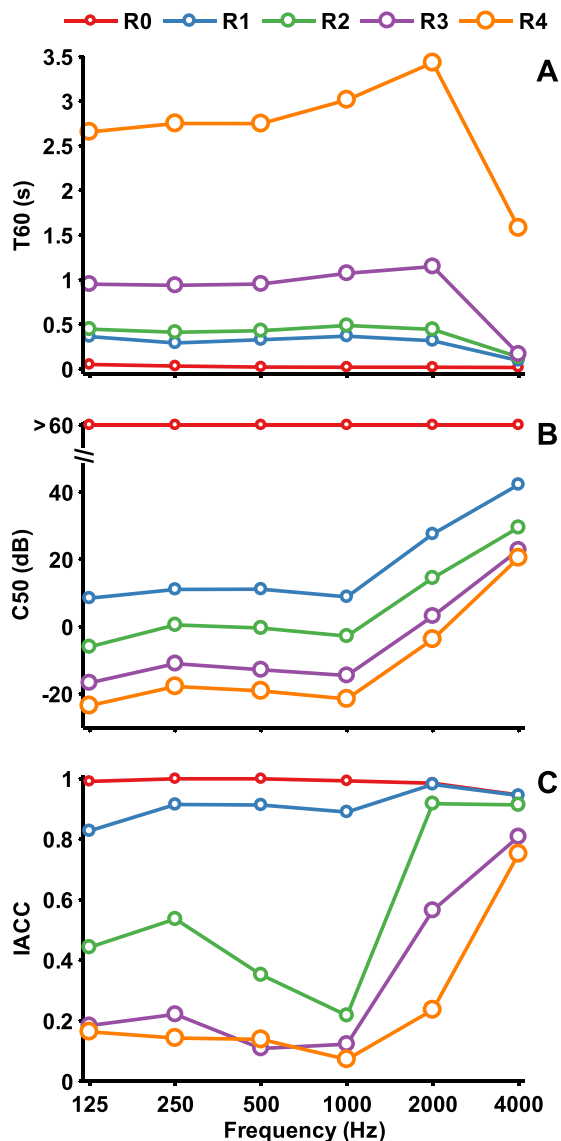


FIG. 1. (Color online) Selected room acoustic parameters as a function of frequency for the five simulated sound fields used in this study (R0–R4). (A) Reverberation time, T60. (B) Clarity index, C50. (C) IACC.

more reverberant rooms produced greater at-the-ear sound levels than the less reverberant rooms.

Within each simulated environment, the speech target was positioned 1.4 m directly in front of the listener. A spatially separated masker was presented on all trials and was positioned 1.4 m from the listener’s position directly opposite the listener’s right ear (90° azimuth angle). This listening configuration is identical to that used in Brandewie and Zahorik (2010).

2. Speech corpus and masker

The speech corpus and masker used in this study were identical to those used in Brandewie and Zahorik (2010). The speech stimuli were from the CRM corpus (Bolía *et al.*, 2000), where each speech sentence in this corpus has the format “Ready (Call Sign) go to (Color) (Number) now.” The corpus has eight talkers (four male and four female), eight

call signs (Charlie, Ringo, Laker, Hopper, Arrow, Tiger, Eagle, Baron), four colors (Blue, Red, White, Green), and eight numbers (1–8). All combinations were used in this study. The masker was broadband continuous Gaussian noise, gated on/off at least 500-ms pre/post the speech signals. Speech and masker signals were convolved with the BRIRs for their appropriate locations relative to the listener (speech at 0 deg and masker at 90 deg) in each of five simulated sound fields (R0–R4). All stimuli were presented over equalized headphones (DT 990 Pro, Beyerdynamic, Germany) at a moderate level [70 dB peak sound pressure level (SPL) at the entrance to the ipsilateral ear].

The length of the speech carrier phrase that preceded the color/number target phrase was manipulated to create two listening conditions that varied the amount of exposure time to the simulated sound field within a given trial of the experiment. In the sentence carrier (SC) condition, two full-length CRM sentences were presented sequentially with ~2.5 s of silence between the sentences. The talker and call sign for the first sentence were selected at random, but the second sentence was always spoken by the same talker as the first sentence and always had the call sign “Baron.” In the no carrier (NC) condition, the target color and number were presented alone, without any carrier phrase (e.g., “Green Three”).

C. Design

Listeners were tested in both the NC and SC conditions. A method of constant stimuli was used in both conditions with nine signal-to-noise ratios (SNRs), ranging from –28 to +4 dB in 4 dB steps. SNR was manipulated by adjusting the gain of the speech target signal prior to convolution with the BRIRs. The masker level was fixed. Target color and number, and the SNR were selected at random for each trial. In the NC condition, the sound field environment was selected at random (equal probability) from trial to trial across a block of trials. This manipulation was designed to minimize any carry-over effects from exposure to a particular sound field from trial to trial. In the SC condition, sound field was held constant across a block of trials. This provided listeners with consistent exposure to the sound field both within (SC phrase) and across trials. Listeners completed at least 270 trials (30 trials per SNR) for a given simulated sound field in both the NC and SC conditions. Not all listeners were tested in all sound fields and conditions.

D. Procedure

The listener was seated in a sound-attenuating booth (Acoustic Systems, Austin, TX—custom double wall) and listened to the headphone-presented stimuli. The listener’s task was to select the appropriate color and number combination using a computer mouse on a graphical interface. Feedback as to whether the response was correct was provided after every trial. All stimulus presentation and response collection was implemented using MATLAB software (Mathworks Inc., Natick, MA).

E. Data analysis

The proportion of correct color/number responses was computed as a function of SNR for individual subjects in both NC and SC for each of the five simulated sound fields (R0–R4). Logistic functions with upper and lower asymptotes were then fit to the data using the psignifit Toolbox (ver. 2.56; Wichmann and Hill, 2001). The lower asymptote, γ , was set to the chance performance level in the CRM task (1/32) for all fits. Threshold, α ; slope, β ; and upper asymptote, $1 - \lambda$, parameters were estimated using a maximum-likelihood procedure (Wichmann and Hill, 2001). Figure 2 displays an example of a fitted function for subject ZAK in the NC condition, R2 sound field. The threshold parameter, α , represents the SNR that corresponds to the midpoint of the function, which will vary slightly across fits depending on the estimated value of λ . In the example fit shown in Fig. 2, $\alpha = -14.92$, which is the SNR that corresponds to $P(C) = 0.4906$. The fitting procedures used by Brandewie and Zahorik (2010) were similar, but did not estimate the upper asymptote parameter, $1 - \lambda$. Wichmann and Hill (2001) have shown that the inclusion of such a parameter that accounts for “lapses” in subject attention to the task results in better estimates of psychometric function threshold and slope.

Goodness of fit was assessed using a deviance statistic, D , which is defined as $2 \log(L_{\max}/L)$, where L_{\max}/L is a likelihood for the saturated model that has as many estimated parameters as data points relative to the best-fitting model with, in this case three estimated parameters. Lower values of D represent better fits to the data. For assessing goodness of fit for logistic functions, deviance is preferable to R^2 used by Brandewie and Zahorik (2010), because it avoids potential problems with interpretation of R^2 for nonlinear data (Spiess and Neumeyer, 2010).

Because the lapse rate parameter, λ , covaries with the α and β parameters of the fitted functions, bounds must be placed on the estimates λ in order to avoid negative impacts

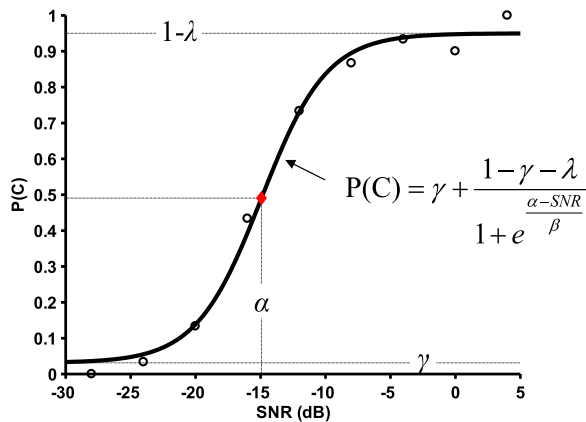


FIG. 2. (Color online) Example psychometric function (logistic), subject ZAK, NC condition, R2 sound field. The deviance statistic, D , for this example was 7.20. The estimated parameters from the fitted psychometric function for this example were: $\alpha = -14.92$, $\beta = 2.51$, $\lambda = 0.05$. The lower asymptote, γ , was held constant at the chance performance level of 1/32 for all fits. Each data point was based on 30 responses, and represents the proportion of correct, $P(C)$, responses.

on the estimation precision of α and β (Wichmann and Hill, 2001). Choices for these bounds were explored by varying the maximum lapse rate value, λ , from 0 (no upper asymptote) to 0.20 and observing the change in the deviance statistic, D , for all data from this study. Figure 3 displays mean D as a function of $\max(\lambda)$. Because a clear knee point in this function may be observed at $\max(\lambda) = 0.05$, estimates of λ were limited to between 0 and 0.05 for all further analyses (upper asymptote, $1 - \lambda$, limited to between 1 and 0.95).

The psignifit toolbox also uses resampling techniques to estimate the sampling distribution of the deviance statistic, D . This distribution can then be compared to the χ^2 distribution because the distribution of D for binomial data is asymptotically distributed as χ^2 . This comparison allows both overdispersion (poor fits to the data), and underdispersion (fits that are better than predicted assuming the data are binomially distributed) to be assessed. Based on these assessments using 9999 iterations in the resampling technique, fits where the p -value of D was found to be 0.975 or greater (overdispersion), or a p -value of 0.025 or less (underdispersion) were excluded. Out of a total of 360 fits, these criteria resulted in 19 fits excluded for overdispersion, and 3 fits excluded for underdispersion. Interested readers are referred to Wichmann and Hill (2001) for a detailed description and analysis of overdispersion and underdispersion pertaining to psychometric function fitting.

Figure 4 displays a matrix of subjects by simulated sound field conditions. Each cell in the matrix indicates whether or not the subject participated in a given condition, and whether fitted psychometric functions were excluded due to either over or under dispersion. Dark shading indicates participation and valid function fits for both NC and SC conditions in a given sound field (a valid fit pair). Light shading indicates at least one excluded fit in either the NC or SC condition for a given sound field (an invalid fit pair). No shading indicates that the subject did not participate in both NC and SC conditions for a given sound field (an incomplete fit pair). The number of valid fit pairs for each simulated sound field is indicated at the bottom of each column. Cells

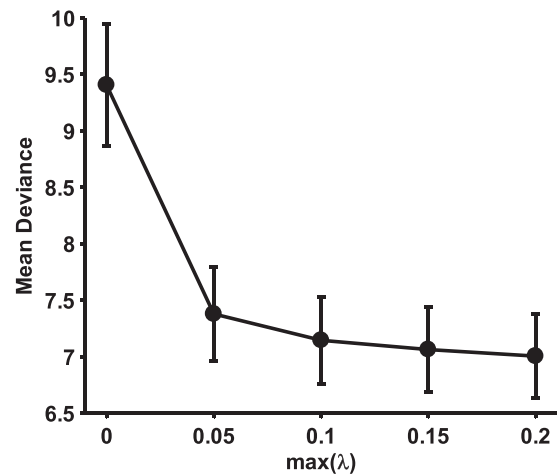


FIG. 3. Mean deviance statistic as a function of the maximum value constraint for the lapse rate parameter, λ , of the fitted psychometric function. Bars indicated 95% confidence intervals.

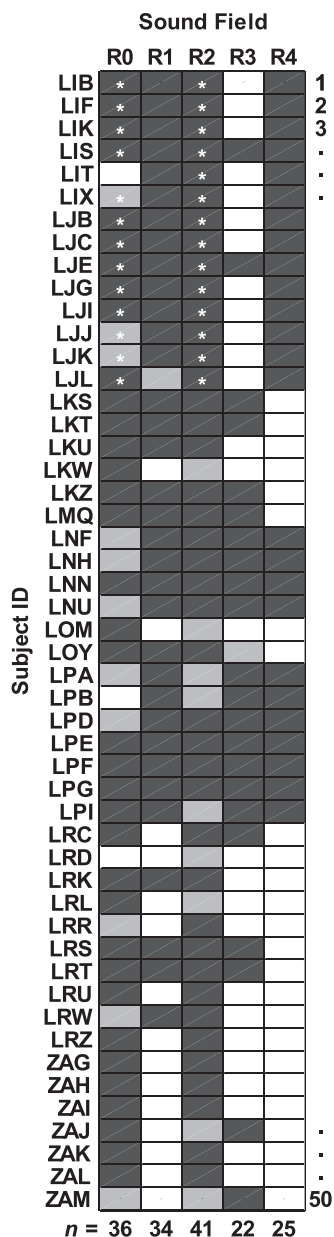


FIG. 4. Sound field by subject participation matrix. Shaded cells indicate subject participation in a given sound field. Not all subjects participated in all sound fields. Dark shading indicates valid psychometric function fits for both SC and NC presentation conditions. Light shading indicates an excluded fit pair (see text for exclusion criteria). Total numbers subjects (n) with valid fit pairs are displayed at the bottom of each column. Cells marked with an asterisk indicate fits based on data from Brandewie and Zahorik (2010).

marked with an asterisk (*) indicate data from Brandewie and Zahorik (2010) that have been refit using the revised procedures described above (e.g., use of deviance statistic to assess goodness-of-fit, and estimation of the upper asymptote, $1 - \lambda$).

Because not all subjects were tested in all simulated sound fields, mixed-model analysis of variance (ANOVA) techniques (cf. Maxwell and Delaney, 2004) were used to evaluate differences in fit parameters and goodness-of-fit metrics across conditions. These techniques are preferable to standard repeated-measures ANOVA techniques in situations such as this, because they do not require all subjects to have been tested in all conditions, and they do not require

assumptions as to the covariance structure between repeated measures (e.g., no sphericity assumptions). Here, inspection of the variance-covariance matrix for each of the dependent variables indicated a compound-symmetric structure. A compound symmetric covariance structure was therefore used for all mixed-model ANOVAs, which were implemented using SPSS software (IBM Corp., Armonk, NY).

III. RESULTS

A. Goodness-of-fit

A goodness-of-fit statistic (deviance) is displayed in Fig. 5 for all valid pairs of psychometric function fits (see Fig. 4) in the NC and SC conditions. Given the fit exclusion criteria that eliminated poor fits to the data due to both overdispersion ($p > 0.975$) and underdispersion ($p < 0.025$), all remaining fits can be assumed to be reasonable approximations to the data ($0.025 \leq p \leq 0.975$). Variability among the fits may be observed, however. A two-way mixed-model ANOVA with factors of listening context condition (NC or SC) and simulated sound field (R0–R4) indicated that although there were differences in the goodness-of-fit across simulated rooms, $F(4,295.150) = 5.174$, $p < 0.001$, there were no differences on average between the goodness-of-fit in the NC and SC conditions, $F(1,267.015) = 0.316$, $p = 0.574$, and differences between NC and SC did not depend on the simulated room, $F(4,266.829) = 1.084$, $p = 0.365$. These results are important for further interpretation of any observed differences between the NC and SC listening conditions, since they suggest that any observed differences cannot be attributed to differences in goodness-of-fit.

The same cannot be said for differences in goodness-of-fit across simulated sound field, however, since statistically significant differences in goodness-of-fit were found for the sound field factor. It should be noted, however, that these

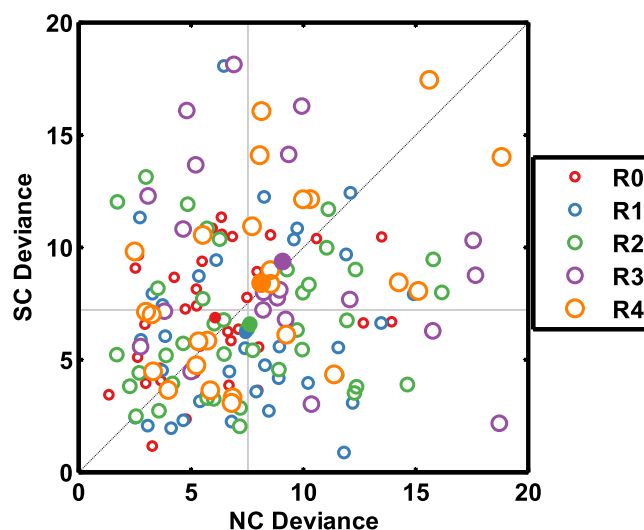


FIG. 5. (Color online) Deviance statistics of the fitted psychometric functions for individual subjects in both NC and SC conditions as a function of simulated sound field (R0–R4). Mean NC and SC deviance values for each simulated sound field are indicated by filled symbols. The gray lines indicated the mean NC and SC deviance values across all sound fields. On average, no consistent differences are observed in goodness-of-fit between the NC and SC conditions.

differences are relatively small and result from slight but statistically significant elevations in deviance for R3, $t(304.048) = 2.739$, $p < 0.001$, and R4, $t(304.597) = 1.750$, $p = 0.010$, relative to the other simulated rooms. Because the deleterious effects of reverberation have long been known, differences in listening performance across the simulated rooms which differ in reverberation is not in itself an emphasis of this study. The interaction of the listening context condition with the sound field simulations is the primary emphasis and, for this relationship, differences in goodness-of-fit were not observed.

B. Psychometric functions and parameter estimates

Figure 6 displays all valid (see Fig. 4) psychometric function fits to individual subject data. The fits are grouped by listening condition (rows) and sound field (columns). The bold functions are averages based on mean parameters (α , β , and λ) for the logistic function (see Fig. 2) for each set of curves. The vertical dashed lines represent the mean threshold parameter, α , for each set. A clear and statistically significant progression of increasing threshold as a function of simulated room may be observed, $F(4,278.398) = 801.146$, $p < 0.001$. This effect is consistent with the well-known inverse relationship between speech intelligibility and the amount of reverberation (Knudsen, 1929).

A perhaps more interesting aspect of the data in Fig. 6 is the apparent interaction between listening context condition and simulated sound field. This interaction is statistically significant, $F(4,258.690) = 8.204$, $p < 0.001$, and results from decreased α in the SC condition relative to the NC condition, the magnitude of which depends on the simulated sound field. To more clearly visualize this effect, differences in the mean threshold parameters are displayed in Fig. 6 both graphically and numerically (in dB). The differences increase as a function of reverberation from R0 to R3, with the largest differences in R2 and R3, but then vanish by R4. Overall, the decreased threshold for SC relative to NC is consistent with the improvements in speech understanding in reverberant environments as reported in previous work (Brandewie and Zahorik, 2010); however, these current results suggest that the effect is strongly room dependent.

Figure 7 displays scatterplots of estimated parameters from the psychometric function fits. In each plot, parameters from the SC condition are plotted against parameters from

the NC condition in order to facilitate a more detailed visual comparison between the conditions than Fig. 6. Sound field simulation is indicated by symbol size and color, as in Fig. 5. Filled symbols indicate the mean parameter estimate for NC and SC conditions.

Estimated threshold parameters, α , are shown in Fig. 7(A). As in Fig. 6, the main effect of increasing threshold with changes in sound field from R0 to R4, and the interaction between listening condition and sound field are clearly evident. The interaction results from a progression of increasingly lower SC thresholds relative to NC from R0 to R3, but then minimal difference in R4.

An additional aspect of the threshold parameter data in Fig. 7(A) that is not clearly evident in Fig. 6 is the relationship between SC and NC α values within a particular simulated sound field. In all cases, there is a strong positive relationship between SC α and NC α , but this relationship is notably less strong in R2 and R3 where the on-average differences between SC α and NC α are the greatest. To more carefully quantify these effects, SC α was regressed on NC α for each simulated sound field. The results of these linear fits are shown in Fig. 7(A) and Table I, where it may be observed that the slopes in R0, R1, and R4 do not differ from 1, but the slopes in R2 and R3 are significantly less than 1. This suggests that the effects of listening context are independent of threshold in R0, R1, and R4, but not for R2 and R3, where context effects appear to increase in magnitude as subjects' performance decreases (elevated threshold parameter, α). In other words, subjects with the lowest thresholds appear to be least affected by listening context, and vice versa: subjects with higher threshold appear to be most strongly affected by listening context.

Figure 7(B) displays a scatterplot of the β parameter, which is related (inversely) to the slope of the psychometric function (see Fig. 2). In general, the β parameter increases as a function of simulated sound field, as indicated by a statistically significant main effect of sound field: $F(4,291.653) = 278.080$, $p < 0.001$. This is similar to the increases in α observed in Fig. 7(A), but here the effect is more strongly driven by increases in β for R3 and R4. For R0, R1, and R2, the β parameters were all very similar, on average. There was also a statistically significant interaction between listening context condition and sound field for the β parameter, $F(4,263.229) = 5.207$, $p < 0.001$. The differences in β parameters for NC versus SC in R3 drive the interaction, as

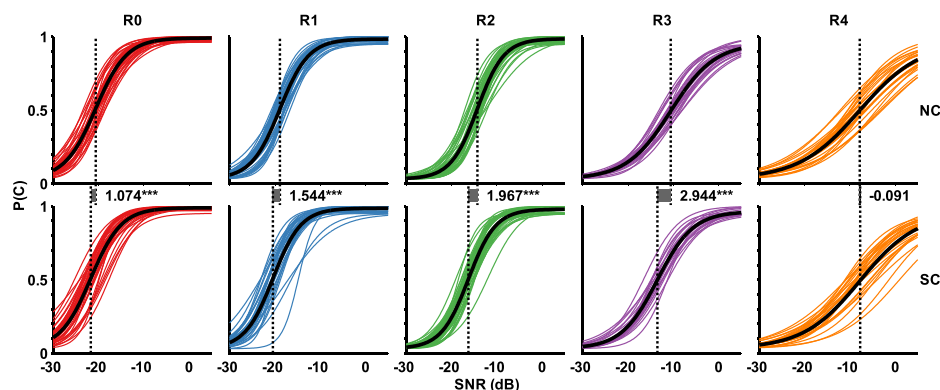


FIG. 6. (Color online) Fitted psychometric functions for individual subjects in both NC (top row) and SC (bottom row) conditions as a function of simulated sound field (R0–R4). For each set of functions, the bold curve represents the average, computed by evaluating the logistic function with mean parameter estimates for the set, and the dashed vertical line indicates the mean threshold parameter. Differences (in dB) between the mean NC and SC thresholds are also indicated for each sound field. “***” indicates $p < 0.001$.

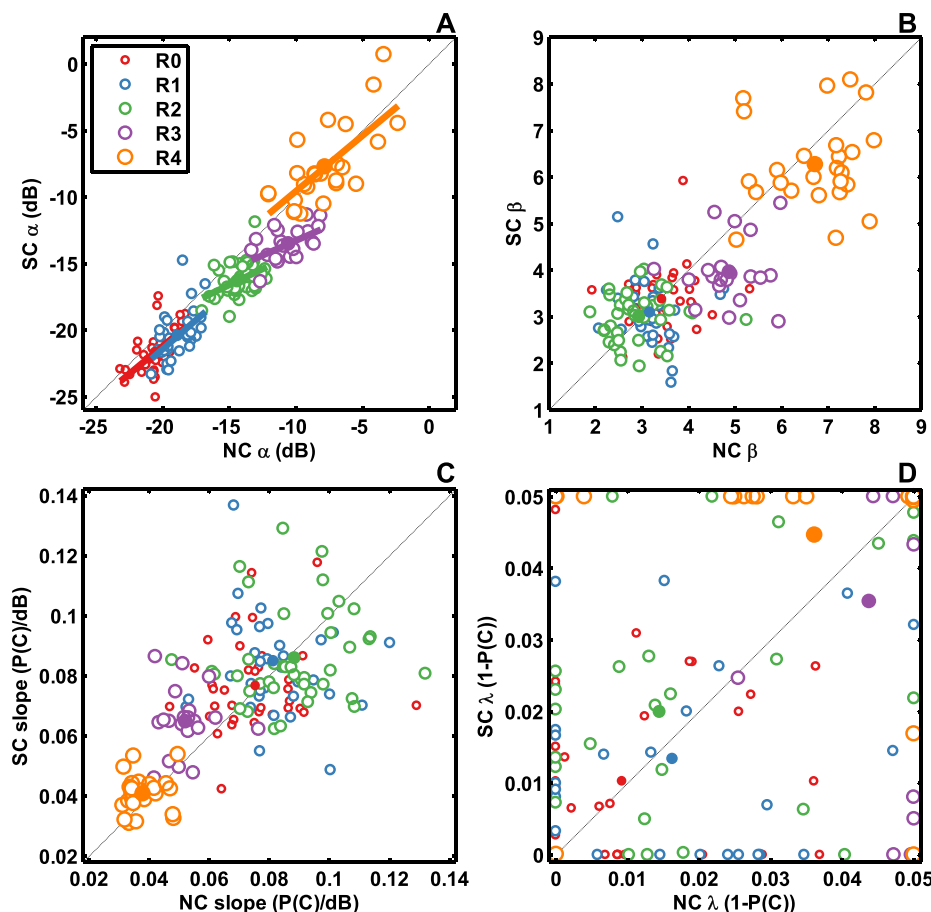


FIG. 7. (Color online) Psychometric function parameter estimates for fitted functions in both NC and SC conditions as a function of simulated sound field (R0–R4). Each data point (open symbol) represents data from an individual subject in a given sound field. Filled symbols indicate parameter means for a given sound field. (A) Threshold parameters, α , with separate linear fits to the data for each sound field shown. (B) β parameter, which controls the slope of the fitted function. (C) Slope as defined by $P(C)/dB$ over the linear portion of the fitted function around threshold. (D) Lapse parameters, λ .

can be seen in Table II, which displays tests of the differences in β between the NC and SC conditions as a function of simulated sound field. These results show that the psychometric functions for speech intelligibility become significantly shallower in the two most reverberant sound fields, but only for R3 are slopes more shallow for NC than for SC listening context conditions. These effects are also apparent visually in Fig. 6.

To more easily interpret the β parameter, the slope of each fitted psychometric function was computed at the relatively linear portion of the function around its midpoint. Results of this transformation are shown in Fig. 7(C). Statistics were not computed on these transformed slope values because the transformation resulted in increased variance heterogeneity. General similarities in the patterns of results between Figs. 7(B) and 7(C) may be observed, however. Shallower function slopes for R3 and R4 relative to R0, R1,

and R2 are evident, as are shallower slopes in R3 for the NC condition relative to the SC condition.

Figure 7(D) displays a scatterplot of the lapse parameter, λ , which controls the upper asymptote of the fitted function (see Fig. 2). Although observable patterns in these parameters are considerably less evident than for either α or β , there was nevertheless a statistically significant main effect of sound field, $F(4,291.012) = 46.023$, $p < 0.001$. This effect is caused by the somewhat elevated λ values for R3 and R4 relative to R0, R1, and R2, although the elevation was only 0.027 on average. There were no appreciable effects of listening context, $F(1,262.304) = 0.166$, $p = 0.684$, or its interaction with sound field, $F(4,262.094) = 2.387$, $p = 0.052$, on the lapse parameter. Lapse parameter effects in this data set therefore appear relatively minor in comparison to the effects of other psychometric function parameters, particularly threshold.

TABLE I. Slopes of best-fitting line relating SC to NC threshold parameters as a function of simulated sound field [see Fig. 7(A)]. Tests of statistical significance (with reference to unity slope) are also shown.

Sound field	Slope (SC α dB / NC α dB)	
R0	0.8318	$t(35) = 0.8401$, $p = 0.2033$
R1	0.8706	$t(33) = 0.5375$, $p = 0.2973$
R2	0.4812	$t(40) = 3.2968$, $p = 0.0010$
R3	0.4434	$t(20) = 3.5569$, $p = 0.0010$
R4	0.8349	$t(24) = 0.9724$, $p = 0.1703$

TABLE II. Within-subjects tests of statistical significance for differences in fitted psychometric function slope parameters, β , between the NC and SC conditions [see Fig. 7(B)] as a function of sound field.

Sound field	
R0	$F(1,35) = 0.136$, $p = 0.714$
R1	$F(1,33) = 0.222$, $p = 0.641$
R2	$F(1,40) = 0.271$, $p = 0.606$
R3	$F(1,22.228) = 24.699$, $p < 0.001$
R4	$F(1,24) = 3.351$, $p = 0.080$

C. SRTs

Because of the inclusion of the lapse parameter, λ , in the fitted psychometric functions, the threshold parameter, α , does not always correspond to the same proportion of correct responses, $P(C)$, across fits. For $\lambda = 0$ and $\gamma = 1/32$ (chance performance), α corresponds to the SNR that produces $P(C) = 0.5156$. For $\lambda = 0.05$ and $\gamma = 1/32$ (chance performance), α corresponds to the SNR that produces $P(C) = 0.4906$. To address this issue, SRTs were computed at the same performance level of $P(C) = 0.50$ for all subjects and conditions, using the estimated parameters from the best-fitting psychometric functions (Fig. 7). Figure 8 displays the mean SRTs (with 95% confidence limits) for each listening context condition (NC and SC) as a function of simulated sound field (R0–R4). Similar to the test results for the α parameters, there was a statistically significant interaction between context condition and sound field, $F(4,258.793) = 9.107$, $p < 0.001$, as well as significant main effects of context condition, $F(1,259.039) = 84.463$, $p < 0.001$, and sound field, $F(4,278.404) = 828.118$, $p < 0.001$.

To aid in the interpretation of the SRT values, estimates of SRT were computed based on the STI. These estimates, which are indicated by the shaded region in Fig. 8, were computed using the following procedure. STI values were first computed from the BRIRs (left ear only) for each sound field in the absence of the noise masker. This was accomplished using methods described in IEC-60268-16 (2003) and Schroeder (1981). The STI values for the reverberant sound fields, R1–R4, were 0.9687, 0.8525, 0.7240, and 0.5642. To convert these values to quantities that can be used to estimate the change in SRT as a function of sound field manipulation, the values from the final “SNR” stage of the STI computation were noted, and used as estimates of the increase in SRT for each of the reverberant sound fields relative to mean SRT in the anechoic sound field (R0). This was done using both the mean R0 SRT from the NC and SC conditions as referents. The shaded region in Fig. 8 represents the results of these estimates based on both the NC and SC R0 referents. Overall, the estimated SRTs based on the STI do a good job of predicting SRT, with the notable exceptions of R1–R3 in the NC condition, where the observed SRTs were worse than predicted. Although these exceptions may be partially due to the fact that the STI was designed to best-predict speech understanding for non-isolated words (IEC-60268-16, 2003), they may also be indicative of disruptions in speech processing that result from inconsistent room acoustic context. Regardless of causal interpretation, it is clear that the magnitude of the listening context effect is strongly room dependent.

D. Corresponding changes in speech intelligibility

To better interpret the magnitude of the listening context effects as a function of sound field, the change in the proportion of target words correctly identified between SC and NC conditions was computed. The computation first determined the SNR that produced $P(C) = 0.70$ in the SC condition, and then determined the $P(C)$ in the NC condition that corresponded to this same SNR. The change in $P(C)$, $\Delta P(C)$, was

then noted for each fit pair (see Fig. 4). The value of $P(C) = 0.70$ was chosen so that comparisons between NC and SC conditions could be made over relatively linear portions of the psychometric functions in each case. Negative values of $\Delta P(C)$ indicated poorer speech understanding in the NC condition relative to the SC condition. All computations were based on the best-fitting psychometric function parameters shown in Fig. 7. Distributions of $\Delta P(C)$ are displayed in Fig. 9 for each of the simulated sound fields. The median $\Delta P(C)$ values are indicated for each distribution, along with results of sign tests, which evaluate differences from $\Delta P(C) = 0$ for each median. In general, the pattern of $\Delta P(C)$ results is similar to that observed for SRT: a modest decrease in performance in R0 for NC presentation versus SC, followed by progressively larger decreases from R1 to R3, but then no change by R4. Of particular interest for this analysis is the magnitude of the change, which for the reverberant rooms progresses from -0.138 to -0.199 (R1–R3). This suggests that the consistency of the listening context can influence speech understanding in moderate reverberation by as much as 20%, on average.

IV. DISCUSSION

The fundamental results from this study are consistent with those from previous works: the consistency of the listening context in reverberant rooms fundamentally alters the perception (Watkins, 2005b) and intelligibility (Brandewie and Zahorik, 2010) of speech. Results from this study extend previous work by demonstrating that the effect of listening context is on average highly room dependent, but quite variable from listener to listener. Relevant aspects of these results and their interpretation are discussed separately in this section.

A. Effect of room

For most of the simulated reverberant rooms tested in this study, prior consistent listening exposure to the reverberant sound field results in a lower SRT compared to situations

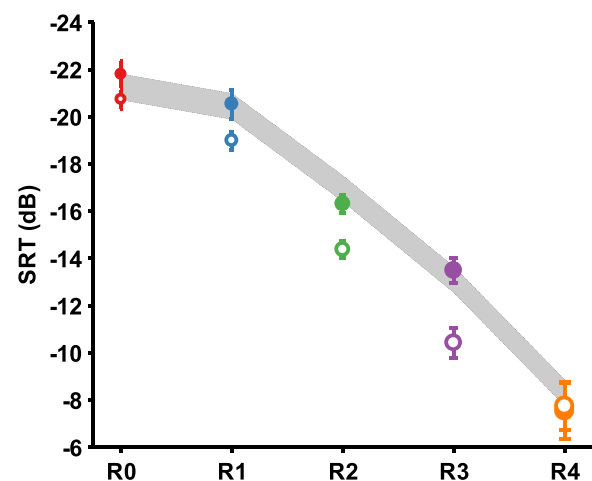


FIG. 8. (Color online) Mean SRT corresponding to 50% correct as a function of simulated room. Bars indicate 95% confidence intervals. Solid symbols are means from the SC condition and open symbols are means from the NC condition. The shaded region represents the estimated SRT based on the STI using R0 SRTs in the NC and SC conditions as anchor points. See the text for details.

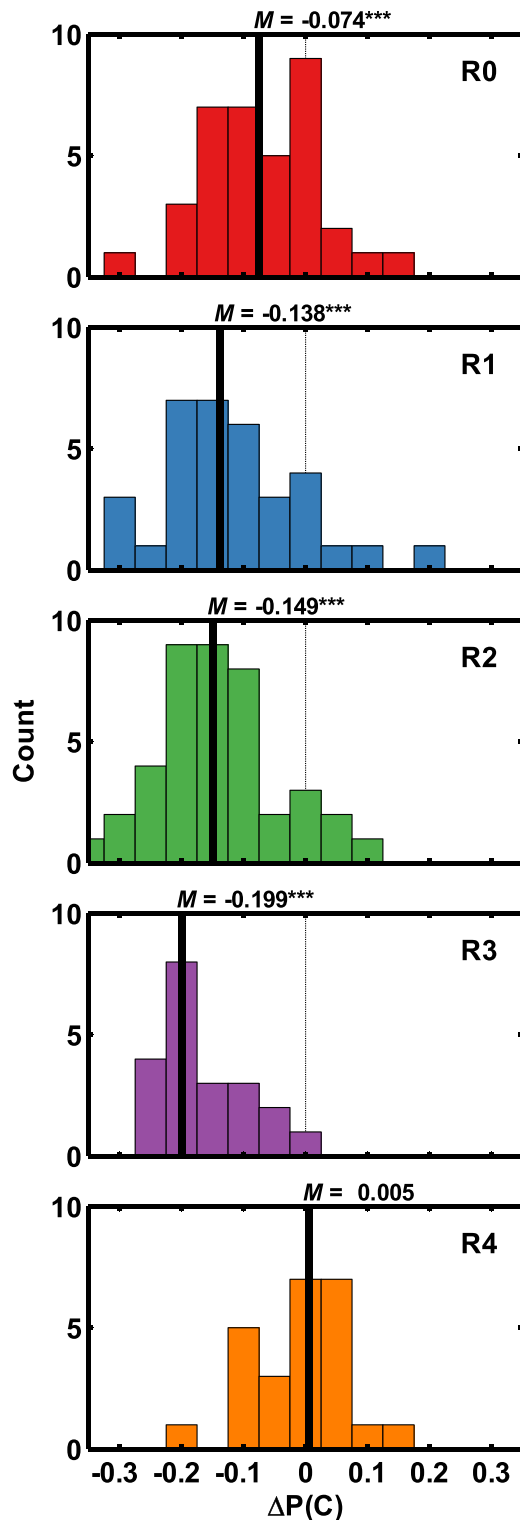


FIG. 9. (Color online) Distributions of changes in the proportions of correct responses, $\Delta P(C)$, in the NC condition relative to $P(C) = 0.7$ for the SC condition as a function of simulated sound field. Negative values indicate poorer performance in NC related to SC. The median (M) $\Delta P(C)$ for each room is indicated numerically and by the bold vertical line. Results of non-parametric sign tests for differences from $\Delta P(C) = 0$ are indicated for each median. “***” indicates $p < 0.001$.

in which prior consistent listening exposure is intentionally limited. The magnitude of the effect appears to scale with the amount of reverberation for moderately reverberant rooms (R1–R3). This result is consistent with Srinivasan and

Zahorik (2014), who demonstrated that when consistent listening exposure is provided for hybrid speech signals (anechoic fine structure, reverberant amplitude envelope), the improvements in the intelligibility were greater for a more-reverberant room ($T60 = 0.7$ s) relative to a less-reverberant room ($T60 = 0.3$ s). It is also consistent with the results from Watkins (2005b) that showed greater perceptual compensation for the effects of reverberation when the listening context is relatively more reverberant (i.e., at a greater distance) and for a larger room. An advantage of the methodology used in the current study is that results from the different simulated rooms can be compared at the same performance level [e.g., $P(C) = 0.50$, as in Fig. 7]. This allows any effects of task difficulty to be removed, since comparisons are made all at the same level of difficulty. This was not the case for Srinivasan and Zahorik (2014), where the two room conditions produced different overall levels of task difficulty, and task difficulty is undefined for the subjective task used by Watkins (2005b).

This pattern of an increasing context effect with increasing reverberation is notably disrupted for the most reverberant room tested, R4, where no context effect was observed. Although the precise cause of this non-effect is uncertain, it may have been due to the amount of “overlap masking,” which is caused by the temporal overlap of sound energy from current speech sounds with the decaying reverberant sound energy from preceding speech sounds (Nabelek *et al.*, 1989). In this study, greater amounts of overlap masking would have occurred in the SC condition than in the NC condition. This could have effectively elevated SRT in SC relative to NC. Because this overlap masking effect is in the opposite direction of the context effect, it may explain why no context effect was observed in R4. Results from Bolt and MacDonald (1949) show that for reverberation time of 3 s, which is similar to that of R4, there is approximately a 20% loss in articulation due to overlap masking. At 70% intelligibility for 32 phonetically balanced (PB) words, a 20% loss in articulation corresponds to roughly a 20% decrease in intelligibility (Kryter, 1962). This decrease is nearly identical in magnitude to the maximum gains intelligibility observed (in R3) due to consistent listening context, which would be consistent with the offsetting effects hypothesis to explain the non-results in R4. Based on estimates from Bolt and MacDonald (1949), the amount of overlap masking in the other less reverberant rooms would have been negligible (at most 5% loss in articulation for R3) and, therefore, likely would have minimally detracted from the observed effects of listening context. It is noted, however, that the predictions of overlap masking effects by Bolt and MacDonald (1949) are based on the perhaps overly simplified assumption that speech signals may be approximated by periodic square pulses. At present, better computational models to explicitly predict the effects of speech overlap masking are not known to exist.

A final and important issue relates to the effects observed in anechoic space (R0). Although Brandewie and Zahorik (2010) did not find a statistically significant context effect in anechoic space and therefore concluded that the context effects reported in that study were specific to

reverberant rooms, results from the current study do show statistically significant differences in performance (threshold parameter, α , and SRT). These differences are likely attributable to known carrier phrase facilitation effects in the SC condition, and do not depend on reverberation per se. For example, Gladstone and Siegenthaler (1971) have shown that the existence of a carrier phrase can improve intelligibility in non-reverberant conditions by about 9% on average. A closer look at Brandewie and Zahorik's (2010) anechoic data shows that, although not statistically significant, there was a mean carrier phrase facilitation effect of 0.84 dB evident at threshold, defined as $P(C) = 0.5156$. The mean R0 effect size from the current data at $P(C) = 0.5156$ is 1.06 dB. This effect size does not differ significantly from Brandewie and Zahorik's (2010) mean result, $t(18.06) = 0.4422$, $p = 0.6636$, but is statistically different from zero, $t(35) = 4.49$, $p < 0.0001$. It is therefore concluded that there likely is a carrier phrase facilitation effect, on the order of 1 dB, and this effect was detectable in the current study due to its larger sample size and corresponding increase in statistical power.

Due to significant carrier phrase facilitation effects in anechoic space, the interpretation of the context effects observed in the reverberant rooms (R1–R3) is somewhat more complicated, as they likely represent a combination of carrier phrase and context effects. If one assumes that the carrier phrase effect observed in anechoic space contributes in the same magnitude to the reverberant room conditions, then the unique contributions of the room context effects would all be reduced by carrier phrase effect. At $P(C) = 0.50$, the smallest room context effect of 1.550 dB observed in R1 (see Table IV) would reduce to ~ 0.4880 dB, which would still be statistically greater than zero, $t(33) = 1.837$, $p = 0.038$. The larger effects observed in R2 and R3 would therefore also still be statistically greater than zero. The non-effect observed in R4 may be a result of overlap masking offsetting not only effects of room context, as discussed above, but also the effect of carrier phrase.

B. Psychometric function slope

Although threshold is by far the most frequently studied parameter of the psychometric function, slope is also critically important, since it determines how rapidly performance can change given changes in the dependent measure. In the current study, psychometric function slopes were found to be relatively constant for R0–R2 at ~ 0.08 P(C)/dB [see Fig. 7(C)], but then decreased for R3 and R4. Additionally, significantly steeper slope was observed in R3's SC condition relative to NC.

In general, this variation in slope is consistent with the results of a recent survey of speech intelligibility studies by MacPherson and Akeroyd (2014) in which the slopes of the 885 psychometric functions from 139 studies were analyzed. For the 18 studies that used the CRM (all conducted in non-reverberant space), the median slope ranged from 0.037 to 0.102 P(C)/dB depending on the type of masking signal and the number of maskers (MacPherson and Akeroyd, 2014). For a single static noise masker, the median slope was 0.101 P(C)/dB with an interquartile range of 0.066–0.136 P(C)/dB (MacPherson and Akeroyd, 2014). The slopes from the

current study for R0–R2 all fell within this range, indicating reasonable consistency across studies.

In addition to slope variations related to the type and number of maskers, MacPherson and Akeroyd (2014) suggest that slope variations may also be related to other factors, including the availability of top-down information, which MacPherson and Akeroyd (2014) discuss in terms of keyword predictability based on the content of the speech materials. They find that steeper slopes are related to higher keyword predictability and therefore more top-down information. This result is consistent with the more general notion that steeper psychometric function slopes are indicative of decreased internal “noise” in the process of perception and response (Swets *et al.*, 1959). It is possible that top-down information contributes to some of the observed differences in slope values, particularly the NC versus SC difference in R3. Watkins (2007a) has suggested that the compensatory effects of speech perception in reverberation observed in his studies may be reflective of a type of high-level perceptual constancy in which top-down information clearly plays a role. The steeper slope in R3 SC relative to NC would be consistent with this idea. The cause of the decreased slopes for R3 and R4 relative to the other sound fields is unknown, but could be related to speech signal degradations caused by greater amounts of reverberation.

C. Variability in effects

In addition to demonstrated average effects of room listening context on speech intelligibility, it is important to recognize the substantial individual variability in these effects. Put simply, different listeners appear to derive different amounts of benefit from room listening context. Although this qualification may appear to be somewhat limiting, it is hardly specific to this particular situation of listening context. Significant individual differences are pervasive in psychoacoustics generally, and in speech intelligibility in particular. In the data presented here, it should be noted that the individual variability in the room context effects is no greater than the variability observed in any of the measures of speech intelligibility. This suggests that the variability in the context effects is, in most instances, driven by the variability in the intelligibility measures in the SC and NC conditions. There is a notable exception to this rule, however. In R2 and R3, the magnitude of the room context effect does appear to be related to performance level in the intelligibility task. Listeners that perform more poorly in the NC task seem to benefit more from the consistent room listening context exposure in the SC condition, as evidenced by the linear fits to the α parameter in R2 and R3 that have slopes less than 1. Although the precise cause for this effect is unknown, it does appear to be specific to the effects of room context, since it is not observed in sound fields that produced little to no context effects (e.g., R0, R1, and R4). The effect of the carrier phrase, as discussed above, also does not seem to produce similar differential benefit for the poorer performing listeners.

D. Enhancement or disruption?

A larger issue relates to the interpretation of the observed room context effects. Does consistent room

listening context lead to enhanced speech understanding, or does inconsistent listening exposure lead to disrupted speech understanding? Although this question may seem paradoxical, it is critical to the ultimate understanding of the mechanisms that may underlie the effect, even if it has minimal impact on the potential functional benefits. At its core, this is a question of what the appropriate baseline is for assessing speech understanding. In previous work (Brandewie and Zahorik, 2010), it was implicitly assumed that the baseline was the condition which minimized prior listening context effects by changing the sound field from trial to trial in the experiment (the NC condition). Consistent listening exposure can then be viewed as a form of enhancement. Of course, situations with inconsistent room listening exposure are rarely, if ever, encountered in the real world. It may therefore be more appropriate to view the situation with consistent listening exposure as the baseline, in which case inconsistent exposure disrupts performance. Because the effects of listening context have been shown to be fast acting, on the order of 1 s (Brandewie and Zahorik, 2013), standard methods of estimating speech understanding that are based on stimulus exposure durations of >1 s (e.g., STI, IEC-60268-16, 2003) can be viewed as adopting this latter view of baseline. This view is consistent with the data shown in Fig. 7, where SRT in the NC conditions is elevated relative to the SC condition and that predicted by the STI. It is also consistent with the interpretation offered by Nielsen and Dau (2010) of the studies by Watkins (2005a,b) and Watkins and Makin (2007a,b).

Regardless of enhancement/disruption interpretation, it is important to note two factors that do not contribute to the differences in performance observed between NC and SC conditions. First, quantifying the effects in terms of changes in SRT guarantees that listeners are performing at the same levels of task difficulty in the two conditions, so differences in task difficulty cannot explain the enhancement/disruption effects. Second, the fact that the lapse rate parameters in the NC and SC conditions are similar suggests that differences in performance are not attributable to differential lapses in attention in one condition relative to the other. This is important, particularly if one adopts the disruption hypothesis, as it suggests that the disruption is perceptual in nature, not attentional. In either case, it is clear that there are functional benefits of brief but consistent prior listening exposure for speech understanding in reverberant sound fields.

E. Potential mechanisms

There is now emerging evidence that the alterations in speech perception and benefits for speech intelligibility resulting from consistent prior listening exposure to reverberant sound fields result from processing of the amplitude envelope of the speech signal (Watkins *et al.*, 2011; Srinivasan and Zahorik, 2014). The results from the present study are generally consistent with this view, since it is evident that as distortion in amplitude envelope caused by the room increases (STI decreases) from R1 to R3, the magnitude of the context effect increases. Exceptions to this pattern are seen in R0, where a small context effect appears

to result from carrier phrase only, and in R4, where overlap masking likely clouds potential context effects. Brandewie and Zahorik (2010) suggested that the context effects observed in that study may have been related to dynamic precedence effect phenomena, in which the auditory system's ability to suppress the contributions of reflected sound is strengthened with repeated exposure. This suggestion was driven in part by the observation that the context effects observed by Brandewie and Zahorik (2010) appeared to require binaural input, since consistent context effects were not observed for monaural input.

Although these two hypotheses regarding the underlying causes of the context effects appear to be quite different, it is possible to re-conceptualize the precedence effect in terms of a pattern-recognition problem in the modulation domain. For example, the simple situation of a click signal followed by a single echo with delay d will produce a modulation spectrum with nulls at frequencies related to d (e.g., $1/2d$, $3/2d$, $5/2d$, etc.). Extraction of the signal and suppression of the echo could be considered as a process that suppresses the nulls in the modulation spectrum due to the echo. Of course more complicated signal patterns with multiple spatially distributed echoes would produce much more complicated modulation patterns that are likely different at the two ears. The fundamental process of signal enhancement and echo suppression could in theory be handled in the modulation domain, however. The fact that neurons in the inferior colliculus (IC) have been implicated in both modulation processing (Rees and Møller, 1983; Rees and Palmer, 1989; Krishna and Semple, 2000) and the precedence effect (Carney and Yin, 1989) lends strength to this hypothesis, as do recent results that demonstrate both monaural and binaural processing of amplitude modulation (AM) signals in reverberation within the IC (Kuwada *et al.*, 2012, 2014). More recently, Slama and Delgutte (2015) have suggested that the processing of AM in the IC may actually involve two processes, one that is either monaural or binaural and serves to enhance AM generally, and then a second binaural process that specifically enhances AM in reverberation. Together these discoveries appear to indicate that the suppression of reflected sound and reverberation has both monaural and binaural aspects that might explain the seeming discrepancies between Watkins' data (Watkins, 2005b), which show context effects for both monaural and binaural listening, and Brandewie and Zahorik's data (2010), which show context effects primarily under binaural listening. Recent preliminary work has demonstrated that human sensitivity to AM in reverberant sound fields is somewhat better than might be predicted based on the acoustical signals reaching the ears (Zahorik *et al.*, 2011; Zahorik *et al.*, 2012), and that this enhancement appears to depend on consistent prior binaural listening exposure to the sound field (Zahorik and Anderson, 2013). Although analogous monaural testing has not been conducted in humans, Linger and colleagues (Lingner *et al.*, 2013) have shown a lack of AM sensitivity enhancement in gerbils, where binaural differences are minimal given the very small head size of the animal. There are also other examples of known AM processing plasticity, such as decreased sensitivity following long term exposure (Tansley

and Suffield, 1983), and short term context effects for AM discrimination (Wakefield and Viemeister, 1990). Clearly, additional work is needed to fully identify the mechanisms that underlie the speech understanding context effects observed when listening to sound in rooms, but explanations based on the precedence effect or modulation processing may not necessarily be incompatible, and the processes may well have both monaural and binaural components.

V. SUMMARY AND CONCLUSIONS

Results from this study confirm past results that demonstrate improved speech intelligibility in a reverberant room when consistent listening exposure to the room is available, relative to situations with inconsistent room exposure. New from this study is the demonstration that the effect of listening exposure is highly room dependent, but also varies from listener to listener. Detailed analysis of psychometric functions for speech intelligibility from 49 normal-hearing listeners quantified individual variability and showed that:

- (1) On average, exposure effects on thresholds were minimal in anechoic space, increased to a maximum of 3 dB (improvement in keyword intelligibility of nearly 20%) in moderate reverberation ($T_{60} = 1$ s), and returned to minimal levels again in high reverberation. The lack of an effect in high reverberation may have resulted from a secondary influence of overlap masking, however.
- (2) For rooms with moderate reverberation ($0.4 \leq T_{60} \leq 1$ s), poorer performing listeners generally showed the greatest effect of room exposure.
- (3) Slopes of the psychometric functions became more shallow, on average, as room reverberation increased. Only in one room (R3) was this effect influenced by the exposure condition.

ACKNOWLEDGMENTS

The authors wish to thank Noah Jacobs, Jeremy Schepers, and Johanna Ohlendorf for their assistance in data collection. Work supported by National Institutes of Health—National Institute on Deafness and Other Communication Disorders (NIH-NIDCD; Grant No. R01DC008168).

Allen, J. B., and Berkley, D. A. (1979). "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.* **65**, 943–950.

ANSI-S3.5-1997 (1997). "Methods for calculation of the speech intelligibility index" (American National Standards Institute, New York).

Beeston, A. V., Brown, G. J., and Watkins, A. J. (2014). "Perceptual compensation for the effects of reverberation on consonant identification: Evidence from studies with monaural stimuli," *J. Acoust. Soc. Am.* **136**, 3072–3084.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). "A speech corpus for multitalter communications research," *J. Acoust. Soc. Am.* **107**, 1065–1066.

Bolt, R. H., and MacDonald, A. D. (1949). "Theory of speech masking by reverberation," *J. Acoust. Soc. Am.* **21**, 577–580.

Brandewie, E., and Zahorik, P. (2010). "Prior listening in rooms improves speech intelligibility," *J. Acoust. Soc. Am.* **128**, 291–299.

Brandewie, E., and Zahorik, P. (2013). "Time course of a perceptual enhancement effect for noise-masked speech in reverberant environments," *J. Acoust. Soc. Am.* **134**, EL265–EL270.

Brown, A. D., Stecker, G. C., and Tollin, D. J. (2015). "The precedence effect in sound localization," *J. of the Assoc. Res. Otolaryngol.* **16**, 1–28.

Carney, L. H., and Yin, T. C. (1989). "Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks: Excitatory and inhibitory components," *J. Neurophysiol.* **62**, 144–161.

Gelfand, S. A., and Hochberg, I. (1976). "Binaural and monaural speech discrimination under reverberation," *Audiology* **15**, 72–84.

Gladstone, V. S., and Siegenthaler, B. M. (1971). "Carrier phrase and speech intelligibility test score," *J. Aud. Res.* **11**, 101–103.

Houtgast, T., and Steeneken, H. J. M. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* **77**, 1069–1077.

IEC-60268-16 (2003). "Sound system equipment—Part 16: Objective rating of speech intelligibility by speech transmission index" (International Electrotechnical Commission, Geneva, Switzerland).

ISO-3382 (1997). "Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters" (International Organization for Standardization, Geneva, Switzerland).

Johnson, J. A., Cox, R. M., and Alexander, G. C. (2010). "Development of APHAB norms for WDRC hearing aids and comparisons with original norms," *Ear Hear.* **31**, 47–55.

Knudsen, V. O. (1929). "The hearing of speech in auditoriums," *J. Acoust. Soc. Am.* **1**, 56–82.

Krishna, B. S., and Semple, M. N. (2000). "Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus," *J. Neurophysiol.* **84**, 255–273.

Kryter, K. D. (1962). "Methods for the calculation and use of the articulation index," *J. Acoust. Soc. Am.* **34**, 1689–1697.

Kuwada, S., Bishop, B., and Kim, D. O. (2012). "Approaches to the study of neural coding of sound source location and sound envelope in real environments," *Front. Neural Circuits* **6**, 1–12.

Kuwada, S., Bishop, B., and Kim, D. O. (2014). "Azimuth and envelope coding in the inferior colliculus of the unanesthetized rabbit: Effect of reverberation and distance," *J. Neurophysiol.* **112**, 1340–1355.

Lingner, A., Kugler, K., Grothe, B., and Wiegand, L. (2013). "Amplitude-modulation detection by gerbils in reverberant sound fields," *Hear. Res.* **302**, 107–112.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (1999). "The precedence effect," *J. Acoust. Soc. Am.* **106**, 1633–1654.

Lochner, J. P. A., and Burger, J. F. (1964). "The influence of reflections on auditorium acoustics," *J. Sound Vib.* **1**, 426–454.

MacPherson, A., and Akeroyd, M. A. (2014). "Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey," *Trends Hear.* **18**, 1–26.

Maxwell, S. E., and Delaney, H. D. (2004). "Designing experiments and analyzing data: A model comparison perspective" (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 763–820.

Moncur, J. P., and Dirks, D. (1967). "Binaural and monaural speech intelligibility in reverberation," *J. Speech Hear. Res.* **10**, 186–195.

Nabelek, A. K., Letowski, T. R., and Tucker, F. M. (1989). "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.* **86**, 1259–1265.

Nabelek, A. K., and Mason, D. (1981). "Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms," *J. Speech, Lang. Hear. Res.* **24**, 375–383.

Nielsen, J. B., and Dau, T. (2010). "Revisiting perceptual compensation for effects of reverberation in speech identification," *J. Acoust. Soc. Am.* **128**, 3088–3094.

Rees, A., and Møller, A. R. (1983). "Responses of neurons in the inferior colliculus of the rat to AM and FM tones," *Hear. Res.* **10**, 301–330.

Rees, A., and Palmer, A. R. (1989). "Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise," *J. Acoust. Soc. Am.* **85**, 1978–1994.

Schroeder, M. R. (1981). "Modulation transfer functions: Definition and measurement," *Acustica* **49**, 179–182.

Slama, M. C., and Delgutte, B. (2015). "Neural coding of sound envelope in reverberant environments," *J. Neurophysiol.* **35**, 4452–4468.

Spiess, A. N., and Neumeyer, N. (2010). "An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach," *BMC Pharmacol.* **10**, 6.

Srinivasan, N. K., and Zahorik, P. (2013). "Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences," *J. Acoust. Soc. Am.* **133**, EL33–EL39.

Srinivasan, N. K., and Zahorik, P. (2014). "Enhancement of speech intelligibility in reverberant rooms: Role of amplitude envelope and temporal fine structure," *J. Acoust. Soc. Am.* **135**, EL239–EL245.

- Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (1959). "Multiple observations of signals in noise," *J. Acoust. Soc. Am.* **31**, 514–521.
- Tansley, B. W., and Suffield, J. B. (1983). "Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation," *J. Acoust. Soc. Am.* **74**, 765–775.
- Wakefield, G. H., and Viemeister, N. F. (1990). "Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise," *J. Acoust. Soc. Am.* **88**, 1367–1373.
- Watkins, A. J. (2005a). "Perceptual compensation for effects of echo and of reverberation on speech identification," *Acta Acust. Acust.* **91**, 892–901.
- Watkins, A. J. (2005b). "Perceptual compensation for effects of reverberation in speech identification," *J. Acoust. Soc. Am.* **118**, 249–262.
- Watkins, A. J., and Makin, S. J. (2007a). "Perceptual compensations for reverberation in speech identification: Effects of single-band, multiple-band and wideband noise contexts," *Acta Acust. Acust.* **93**, 403–410.
- Watkins, A. J., and Makin, S. J. (2007b). "Steady-spectrum contexts and perceptual compensation for reverberation in speech identification," *J. Acoust. Soc. Am.* **121**, 257–266.
- Watkins, A. J., Raimond, A. P., and Makin, S. J. (2011). "Temporal-envelope constancy of speech in rooms and the perceptual weighting of frequency bands," *J. Acoust. Soc. Am.* **130**, 2777–2788.
- Wichmann, F. A., and Hill, N. J. (2001). "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept. Psychophys.* **63**, 1293–1313.
- Zahorik, P. (2009). "Perceptually relevant parameters for virtual listening simulation of small room acoustics," *J. Acoust. Soc. Am.* **126**, 776–791.
- Zahorik, P., and Anderson, P. W. (2013). "Amplitude modulation detection by human listeners in reverberant sound fields: Effects of prior listening exposure," *Proc. Meet. Acoust.* **19**, 050139.
- Zahorik, P., Kim, D. O., Kuwada, S., Anderson, P. W., Brandewie, E., Collecchia, R., and Srinivasan, N. (2012). "Amplitude modulation detection by human listeners in reverberant sound fields: Carrier bandwidth effects and binaural versus monaural comparison," *Proc. Meet. Acoust.* **15**, 050002.
- Zahorik, P., Kim, D. O., Kuwada, S., Anderson, P. W., Brandewie, E., and Srinivasan, N. (2011). "Amplitude modulation detection by human listeners in sound fields," *Proc. Meet. Acoust.* **12**, 050005.